Two-microphone spatial filtering provides speech reception benefits for cochlear implant users in difficult acoustic environments

Raymond L. Goldsworthy[a)]

Research & Development, Sensimetrics Corporation, 14 Summer Street, Suite 305, Malden, Massachusetts 02148

Lorraine A. Delhorne,[b)] Joseph G. Desloge,[c)] and Louis D. Braida

Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 36-791, Cambridge, Massachusetts 02139

(Received 21 September 2013; revised 10 June 2014; accepted 19 June 2014)

This article introduces and provides an assessment of a spatial-filtering algorithm based on two closely-spaced (~1 cm) microphones in a behind-the-ear shell. The evaluated spatial-filtering algorithm used fast (~10 ms) temporal-spectral analysis to determine the location of incoming sounds and to enhance sounds arriving from straight ahead of the listener. Speech reception thresholds (SRTs) were measured for eight cochlear implant (CI) users using consonant and vowel materials under three processing conditions: An omni-directional response, a dipole-directional response, and the spatial-filtering algorithm. The background noise condition used three simultaneous time-reversed speech signals as interferers located at 90°, 180°, and 270°. Results indicated that the spatial-filtering algorithm can provide speech reception benefits of 5.8 to 10.7 dB SRT compared to an omni-directional response in a reverberant room with multiple noise sources. Given the observed SRT benefits, coupled with an efficient design, the proposed algorithm is promising as a CI noise-reduction solution.

 [http://dx.doi.org/10.1121/1.4887453]

PACS number(s): 43.71.Ky, 43.72.Dv, 43.66.Ts, 43.60.Fg [VMR] Pages: 867–876

I. INTRODUCTION

The aim of this study was to evaluate a spatial-filtering algorithm designed to improve speech reception in noise for cochlear implant (CI) users. Many factors contribute to poor speech reception in noise for CI users including reduced number of surviving nerve fibers (Morita *et al.*, 2004), limited electrical dynamic range that is less than 20 dB (Zeng *et al.*, 2002), reduced spectral resolution resulting from the limited number and location of implanted electrodes (Hanekom and Shannon, 1998; Henry and Turner, 2003; Hughes and Goulson, 2011), reduced temporal resolution associated with the carrier that modulates the electric pulse train (Muchnik, 1994; Fu, 2002), and loss of information about stimulus fine structure (Smith *et al.*, 2002). Consequently, compared to normal-hearing listeners, CI users require an increase of 5 to 17 dB in the signal-to-noise ratio (SNR) when tested using stationary speech-shaped noise (Hochberg *et al.*, 1992; Fu *et al.*, 1998). When tested with modulated noise, CI users require at least a 20 dB increase in SNR to achieve comparable performance as normal-hearing listeners (Nelson *et al.*, 2003; Goldsworthy *et al.*, 2013).

Hu and Loizou (2008) provided a clear proof-of-concept for a promising class of CI noise reduction strategies. They demonstrated that speech reception in noise could be restored to performance levels in quiet simply by turning off CI filters whenever the SNR dropped below 0 dB SNR within that filter. They found that this method improved speech reception for both speech-shaped noise and babble using broadband SNRs between 0 and 10 dB. This elegant proof demonstrated that CI speech reception can be improved in noisy situations; however, this proof required foreknowledge of SNRs within individual filters. The present article evaluates a spatial-filtering algorithm for CI users (Goldsworthy *et al.*, 2009) dubbed "Fennec" after the African desert fox with exceptional hearing, which builds upon these observations. While the proof-of-concept provided by Hu and Loizou required foreknowledge of SNR characteristics, the Fennec strategy achieves the same objective using real-time acoustic analysis. Specifically, this strategy uses multiple microphones to identify and to preserve target-dominated components while attenuating noise-dominated components.

In practice, a number of noise reduction strategies have been evaluated for use with CIs. Noise reduction strategies that use a single microphone are designed to exploit temporal and/or spectral differences between speech and noise in order to attenuate time-frequency regions where SNR is poor. While these methods have not improved speech

[a)]Author to whom correspondence should be addressed. Electronic mail: raygold@sens.com

[b)]Current address: Department of Biomedical Engineering, Boston University, Hearing Research Center, Room 412B, 44 Cummington Mall, Boston, MA 02215.

[c)]Also at: Research & Development, Sensimetrics Corporation, 14 Summer Street, Suite 305, Malden, MA 02148.

reception in noise for normal-hearing listeners (Lim and Oppenheim, 1979), they have provided benefits of 4 to 5 dB for CI users (Hochberg *et al.*, 1992; Weiss, 1993; Yang and Fu, 2005; Goldsworthy, 2005; Hu and Loizou, 2010; Dawson *et al.*, 2011). While such benefits are promising, performance gains are generally limited to situations when the noise source has a constant, well-defined, acoustic spectrum such as an air-conditioner.

Multiple-microphone strategies are inherently more powerful than single-microphone strategies since they incorporate spatial information in addition to any temporal or spectral differences between sounds. The output SNR of a multiple-microphone array can be increased using a linear, time-invariant, combination of the microphone signals. Such fixed-processing beamformers have been demonstrated as effective noise reduction solutions for CIs (Chung *et al.*, 2004, 2006; Chung and Zeng, 2009).

Null-steering beamforming is a class of multiple-microphone noise reduction (which includes "generalized sidelobe cancellers" and "linearly-constrained adaptive beamforming") which adaptively updates the signal processing to steer spatial nulls in order to cancel noise sources in the acoustic environment (Frost, 1972; Widrow *et al.*, 1975; Griffiths and Jim, 1982). The BEAM™ strategy (Spriet *et al.*, 2007; Hersbach *et al.*, 2012) that has been implemented on Cochlear Corporation devices is an example of a null-steering beamformer that has been developed for CI use.

Relative to an omni-directional reference, null-steering beamforming based on unilateral, closely-spaced microphones, have improved speech reception thresholds (SRTs) by up to 7 to 16 dB for CI users in laboratory listening conditions (Wouters and Vanden Berghe, 2001; Spriet *et al.*, 2007). However, performance deteriorates as acoustic conditions degrade due to complicating factors such as reverberation and moving or multiple noise sources (Greenberg and Zurek, 1992; van Hoesel and Clark, 1995; Hamacher *et al.*, 1997; Wouters and Vanden Berghe, 2001).

The spatial-filtering algorithm evaluated in the present study augments very simple directional-microphone beamforming with adaptive signal processing that selectively attenuates signal components that appear to be dominated by noise. The emergence of this class of spatial filtering was inspired by physiological models of binaural hearing (Jeffress, 1948; Colburn, 1994). Kollmeier and colleagues (Kollmeier *et al.*, 1993; Kollmeier and Koch, 1994) developed and evaluated a spatial-filtering algorithm, using binaurally-situated microphones, which provided an SRT gain of roughly 2 dB relative to an omni-directional reference in background noise. Margo *et al.* (1997) evaluated a similar strategy using a real-time device and observed mixed results. Specifically, in a sound-proof room with straight-ahead target speech and noise at 90°, 3 of 11 listeners showed worse performance with the processing, while the remaining 8 had speech-reception improvements ranging from 20 to 77 percentage points. More recently, binaural noise reduction has been shown to improve performance for CI users (Goldsworthy, 2005; Kokkinakis and Loizou, 2008) with speech reception benefits in noise as large as 60% on keyword recognition and benefits of 14 dB in SRTs.

Goldsworthy *et al.* (2009) transitioned this style of spatial filtering to a configuration with two closely-spaced (~2 cm) microphones mounted in a behind-the-ear (BTE) shell; they argued that this approach was more robust to the effects of reverberation and demonstrated speech reception benefits for CI users. Yousefian and Loizou developed a similar approach based on the coherence between microphones and measured benefits of 30% to 60% in word recognition for normal-hearing listeners tested on IEEE sentences in single- and multiple-noise source scenarios (Yousefian *et al.*, 2010; Yousefian and Loizou, 2012, 2013). In their 2013 study, Yousefian and Loizou found that their algorithm, compared to a fixed directional algorithm, provided benefits of 5 to 10 dB SRT for CI users when tested in a nearly anechoic environment. They found that this benefit was affected by reverberation with benefits decreasing to 4 to 7 dB SRT and to 1 to 2 dB SRT when tested in rooms with reverberation times (T_{60}) of 220 and 465 ms, respectively.

Hersbach *et al.* (2013) investigated the combination of null-steering beamforming with a spatial-filtering algorithm similar to the approach of Goldsworthy (2009) and of Yousefian and Loizou (2012, 2013). Specifically, they used a null-steering beamformer to approximate the instantaneous SNR and then selectively attenuate spectral components that statistically appeared to be dominated by noise. They found that this combined approach yielded an additional benefit of 4.6 dB SRT compared to null-steering beamforming when testing CI users in a sound-treated, low-reverberation (the T_{60} was not specified), multiple-noise-source environment.

Given these initial and promising results of spatial-filtering for improving speech reception in noise for CI users, the present study evaluates the algorithm introduced by Goldsworthy *et al.* (2009). This approach estimates the direction-of-arrival for each spectral component within a short-time Fourier transform (STFT); components that have phase characteristics indicating straight-ahead direction-of-arrival are preserved, while components are increasingly attenuated as the phase signature indicates rear direction-of-arrival.

This approach is based on direct analysis of phase differences, rather than the use of coherence as suggested by Yousefian and Loizou (2013). This distinction can yield performance differences, particularly in more complex environments where the target speech is degraded by reverberation and by multiple, simultaneous, noise sources. Hersbach *et al.* (2013) suggested it was this dependence on coherence which caused the Yousefian and Loizou (2013) algorithm to deteriorate in reverberant environments. Specifically, speech reception benefits provided by the coherence-based algorithm decreased from 5 to 10 dB in an anechoic environment to 0 to 2 dB in a moderately reverberant room ($T_{60} = 465$ ms). Hersbach *et al.* (2013) suggested an alternate approach, using null-steering beamforming as a front-end to a secondary post-filter using spectral-attenuation of low-SNR components. As they only evaluated this method in a sound-treated, low-reverberation environment, it is unknown the extent to which that approach is robust to the detrimental effects of higher levels of reverberation. Since performance in reverberation is a highly relevant metric of success for CI noise reduction, the present study evaluates the Fennec spatial-filtering algorithm in a reverberant ($T_{60} = 350$ ms), multiple-noise source, condition using consonant- and vowel-identification measures.

II. METHODS

A. Subjects

Eight adult CI users participated in this study with relevant information summarized in Table I. All subjects except S8 had previously participated in at least one other psychoacoustic and speech experiment in our laboratory. Subjects provided informed consent on their first visit to the laboratory and were paid for their participation in the study.

At the time of testing, the subjects ranged in age from 18 to 66 yrs (mean = 49.5 yrs). Five of the eight subjects (S1, S2, S3, S5, and S6) reported that the cause of their hearing loss was either genetic or unknown. In these cases the loss was typically diagnosed at birth or early childhood and progressed over time. Subject S7 had normal hearing until the age of 6 when she contracted mumps which resulted in a hearing loss that progressed over time. Subject S4 had normal hearing until the age of 32 when he lost his hearing gradually due to ototoxicity. Subject S8 was diagnosed with a minimal hearing loss in grade school, which progressed slowly until she needed a hearing aid at the age of 45 yrs. Ten years later she suddenly lost all useable hearing and received an implant.

The age at implantation ranged from less than 2 to 50 yrs (mean = 35) and duration of implant use at the time of testing ranged from 4 to 16 yrs. Three subjects had worn their implants for 10 yrs or more, 4 subjects between 5 and 9 yrs, and 1 subject for less than 5 yrs. The subjects used a variety of implant sound processors, including the Nucleus Freedom, Nucleus 5, Nucleus 3G, Auria 4, and the Harmony.

All subjects were tested monaurally. Subjects S1, S5, and S8 were bilateral implant users who used their better ear for this experiment. For S1 and S5 the better ear was chosen based on the results of previous psychoacoustic tests. In both cases the better ear was the ear that was implanted first and was also the one that the subject favored. S8 did not participate in previous studies in our lab and we chose the ear based on her preference.

B. Target and noise materials

Speech reception in noise was measured separately for two target stimulus sets: Consonant and vowel. The consonant stimulus set was drawn from speech recordings collected by Shannon *et al.* (1999) and consisted of 20 monosyllables in /a/-C-/a/ format for 20 values of C = /b d g p t k m n N l r f v T ð s z Σ tΣ δZ Z φ ʍ η/. Utterances of each of the 20 syllables from 5 female and 5 male talkers were used, yielding a total set of 200 consonant tokens. The vowel stimulus set was drawn from speech recordings collected by Hillenbrand *et al.* (1995) and consisted of ten monophthongs (/i I ε æ u ʊ a ɔ ʌ ɝ/) and two diphthongs (/əʊ eI/), presented in /h/-V-/d/ context (heed, hid, head, had, who would, hood, hod, hud, hawed, heard, hoed, hayed). Utterances of each of the 12 syllables from 5 female and 5 male talkers were used, yielding 120 vowel tokens. The consonant and vowel databases were originally digitized at sampling rates of 44 100 and 16 000 Hz, respectively; vowel materials were resampled to 44 100 Hz and all processing was implemented at a sample rate of 44 100 Hz.

The noise stimuli consisted of time-reversed speech clips formed from recordings of IEEE-sentences (IEEE, 1969) made at House Research Institute. Noise material was randomly selected from this database. For the experimental conditions, the duration of the selected noise material was 2 s plus the duration of the target speech material. This preceding 2 s allowed the convolved reverberation to build for the target plus noise portion of the stimulus. For stimulus presentation, the initial 2 s were not played to the listeners and the SNR did not include the portion where the target material was absent.

C. Stimulus generation and presentation

All stimuli were pre-processed prior to being delivered to the subject. The pre-processing consisted of two steps.

In the first step, one target and three equal-level noise stimuli were convolved with two-channel source-to-microphone impulse responses measured for two microphones positioned 1 cm apart in a BTE shell worn on the left ear of a KEMAR manikin (Knowles Corp., Itasca, IL) (Burkhard and Sachs, 1975). These impulse responses were measured using Maximum Length Sequences (Rife and Vanderkooy, 1989; Vanderkooy, 1994) in a mildly-reverberant laboratory space

TABLE I. Subject information

Subject	Sex	Ear tested	Etiology	Age at onset of hearing loss/deafness	Age at implantation (years)	Age at time of testing	Length of implant use (years)	Implant processor
S1	F	R	Unknown	Birth-progressive	15	19	4	Cochlear Freedom
S2	F	R	Unknown	Birth	21 months	18	16	Nucleus Freedom
S3	F	L	Unknown	Birth-progressive	50	56	6	AB AURIA
S4	M	L	Ototoxicity (Gentamycin)	32-progressive	47	56	9	Cochlear Nucleus 5
S5	M	R	Genetic disorder	Birth-progressive	53	63	10	Cochlear Esprit 3G
S6	F	R	Genetic disorder	Birth-progressive	60	66	6	AB AURIA
S7	F	R	Mumps	diagnosed at age 6 progressive	48	53	5	AB Harmony
S8	F	R	Unknown	45 - progressive	55	65	10	AB Harmony

($4.7 \times 7.4 \times 3.0$ m with a 60-dB reverberation-decay time of 350 ms). Since these impulse responses encompass the actual source-to-microphone propagation information from the measurement room (including the head/torso effects and actual room reverberation), this technique allows a large variety of realistic test stimuli to be synthesized in an efficient manner.

The target impulse response was for a source positioned 0.75 m at 0° directly in front of the KEMAR manikin, while the three noise impulse responses were for sources positioned 0.75 m at 90° to the left of, 180° behind, and −90° to the right of the KEMAR manikin. The target level was scaled for each subject to achieve a comfortable rear-microphone listening level. The three noises were summed and the result was scaled to achieve a specific SNR at the rear microphone. Due to the closely-spaced nature of the microphones, target and noise levels at the front microphone were essentially identical to those at the rear-microphone. The scaled target and noise were then summed to generate a simulated target-plus-noise CI input.

In the second step, the two simulated microphone input signals were processed for noise reduction to yield a single signal for presentation to the listener. Three processing options were considered: Omni-directional, dipole-directional, and Fennec.

For *Omni* processing, the rear BTE microphone signal was presented to the subject. In the free-field (i.e., not mounted on the head), the directional response of this processing would be identical from all directions. Mounted on the head, it yielded no improvements in the target SNR other than those achieved naturally by head-shadow, etc.

For *Dipole* processing, the two BTE microphone signals were combined to yield a dipole directional-microphone response and presented to the subject. The upper branch of the signal processing in Fig. 1 depicts the processing that was used to generate the dipole response. Specifically, the STFT—based on a 256-point analysis window (11.6 ms) and 50%-block overlap—was taken of the two microphone signals. The STFT of the rear microphone signal was then subtracted from the STFT of the front microphone signal and frequency-dependent compensation (capped at a maximum of 18 dB for frequencies below 690 Hz in order to limit noise amplification) was applied to counteract the low-frequency attenuation that resulted from the subtraction (Stadler and Rabinowitz, 1993). The inverse STFT was then taken to yield the processed output. (Note that the phase-based attenuation processing shown in Fig. 1 was part of Fennec processing described below and is not used for pure dipole processing.)

In the free-field (i.e., when not head-worn), the directional response of this processing was the "figure-8" pattern shown as Directional Response A of Fig. 1 in which lateral sources from ±90° are attenuated while sources from the front and back are preserved. Although this response did change when mounted on the head, this processing still yielded directional SNR improvements relative to the omni-directional processing.

For *Fennec* processing, as shown in the schematic from Fig. 1, the two BTE microphone signals were used to form a dipole response as described above, and then an additional, phase-based attenuation term was generated from the two microphone signals and applied to the dipole signal in order to further reduce noise. The phase-based attenuation term was calculated in the following manner. First, the time-dependent cross-spectral power density, $S_{fb}[n,k]$, for the front and back signals was calculated for each STFT frequency bin using

$$S_{fb}[n,k] = \alpha * S_{fb}[n-1,k] + (1-\alpha) * STFT_f[n,k] * conj(STFT_b[n,k]), \quad (1)$$

where $STFT_f[n,k]$ and $STFT_b[n,k]$ were the front- and back-microphone STFTs, respectively, n was the STFT time index, k was the STFT frequency bin, conj() was the complex-conjugate operator, and α was the parameter of a first-order infinite-impulse response filter used to smooth the estimate (we used $\alpha = 0.56$, which yielded a filter time constant of 10 ms).

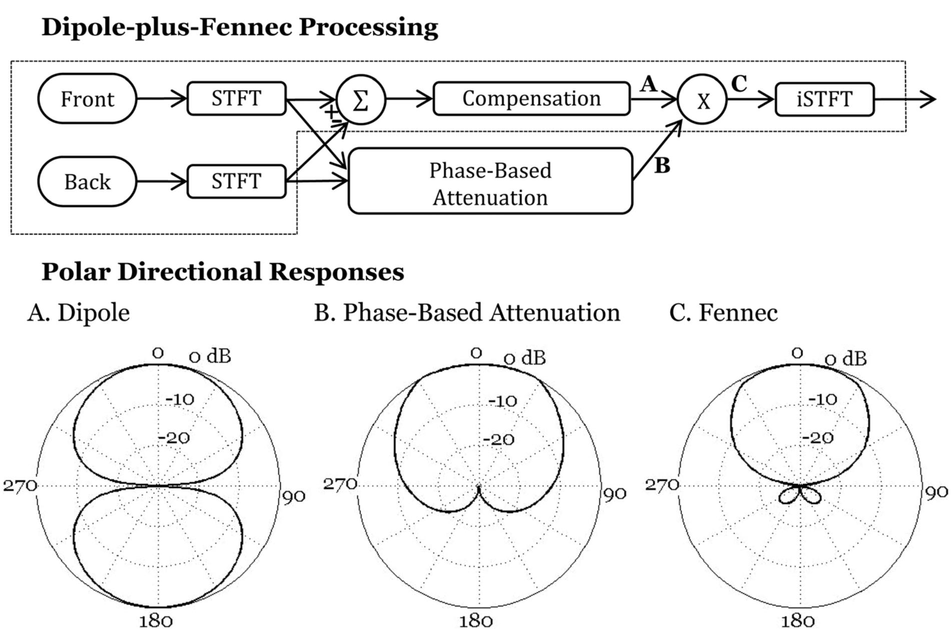

FIG. 1. Dipole-plus-Fennec processing schematic and polar directional responses at three stages within the processing indicated by A, B, and C. The dashed line encompasses the dipole components of the processing schematic. The phase-based attenuation indicates additional time/frequency attenuation for the Fennec strategy.

The phase of $S_{fb}[n,k]$ was then used to estimate the direction-of-arrival that dominated the content of time/frequency cell [n, k] of the STFT. Specifically, based on free-field acoustics and assuming that a single directional source from azimuth location θ (0° = straight ahead) accounted for all energy in the cell [n, k], then

$$\angle\left(S_{fb}[n,k]\right) = \frac{2\pi k d}{Nc}\cos(\theta), \tag{2}$$

where $N = 256$ was the STFT block size, $d = 0.01$ m was the microphone separation, and $c = 345$ m/s was the velocity of sound. By inverting this equation, it was possible to estimate the angle-of-incidence, $\hat{\theta}$, for a single free-field source what would have given rise to the observed phase.

This estimated angle of incidence was then used to calculate an attenuation factor that was applied to each time/frequency cell of the dipole STFT,

$$A[n,k] = \min\left(0, \frac{A_{180}(\hat{\theta} - \beta)}{180 - \beta}\right), \tag{3}$$

where A_{180} was the desired attenuation in dB at 180° and β was the angle-of-incidence below which the attenuation is 0 dB. For the implementation used in this study, A_{180} and β were set to 30 dB and 30°, respectively, to yield the desired attenuation polar pattern shown as Polar Directional Response B of Fig. 1. This attenuation term was calculated for each STFT time/frequency cell of the input signals and was applied to the dipole STFT prior to signal reconstruction via the inverse STFT. The theoretical, free-field Fennec directional response for a single source, combining the dipole and the phase-based-attenuation, is shown as Polar Directional Response C of Fig. 1.

Each pre-processed stimulus signal was presented to the listener through Sennheiser HD580 circumaural headphones with the speaker portion of the headphone placed directly over the CI microphone.

D. Familiarization procedure

Prior to the main experimental phase of this study, each listener completed a familiarization process designed to provide exposure to the test procedure, the test materials, and the three processing schemes (Omni, Dipole, and Fennec). Specifically, for each processing scheme, subjects were asked to identify stimuli presented in three conditions that increased in difficulty: (1) Stimulus identification in quiet with correct-answer feedback, (2) stimulus identification in noise (SNR = 0 dB) with correct-answer feedback, and (3) stimulus identification in noise (SNR = 0 dB) without correct-answer feedback. The three conditions were tested in the same order for each subject, but the test order of the processing schemes within each condition was randomized. For each condition/processing-scheme combination, both the entire consonant target stimulus set (200 tokens) and entire vowel target stimulus set (120 tokens) were selected in a randomized order and all tokens within the selected set were presented in random order without replacement.

E. Speech reception testing

Once a subject completed the familiarization procedure, the three processing strategies (Omni, Dipole, and Fennec) were evaluated by measuring SRTs for both the consonant and vowel databases separately. While it is less common to use monosyllabic words, or closed-set phoneme materials, within adaptive SRT procedures, previous studies have demonstrated the efficacy of such methods (Tecca and Binnie, 1982; Mackie and Dermody, 1986; Wall *et al.*, 1989; Liu and Eddins, 2012). Three separate measures of SRT were obtained for each processing strategy and each target stimulus set (consonant and vowel) by using three adaptive rules (described below) that converged to 70.7%, 50%, and 29.3% correct speech reception. This yielded a total of 18 test conditions (3 processing strategies × 2 target stimulus sets × 3 SRT measures). The testing order for the processing strategies and target stimulus sets was randomized, but the SRT measures were always made in decreasing percent-correct order (first 70.7%, then 50%, and finally 29.3%) for each processing-strategy/target-stimulus-set combination.

SRTs were measured for each processing scheme for both consonant and vowel materials. For a given run, speech tokens were randomly selected without replacement from the target-speech stimulus set (200 tokens for the consonants and 120 tokens for the vowels). The target speech was always presented at the same level (pre-determined for each listener as described above to be a comfortable listening level). For the first presentation in the measurement process, the noise was scaled to achieve an SNR of 12 dB. The SNRs of the remaining presentations were then adjusted up or down using 3 dB steps until the third reversal, then 2 dB steps until the sixth reversal, and then 1 dB steps thereafter depending upon the subjects correct or incorrect response and the desired SRT convergence criterion. Specifically, the 70.7% correct SRT was estimated using a 1-up, 2-down rule, the 50% correct SRT was estimated using a 1-up, 1-down rule, and the 29.3% correct SNR was estimated using a 2-up, 1-down rule (Levitt, 1971). The final estimated SRT was calculated as the average SNR over the final 180 tokens for the consonants or final 100 tokens for the vowels.

After completion of all 18 test conditions, a subset of 6 conditions was repeated as a validity check. Specifically, the conditions measuring the 50% correct SRT for all six combinations of processing strategies and target-speech materials were repeated for all subjects with the following exceptions. Subject S5 could not achieve 50% correct performance, and so the 29.3% correct SRT conditions were repeated instead. Subject S2 was unavailable to complete the validation check for any conditions. Subject S4 was unavailable to complete the validation check for the consonant materials with dipole processing.

III. RESULTS

Figures 2 and 3 summarize the measured SRTs for each subject on the 18 conditions tested. Note that Subjects S3

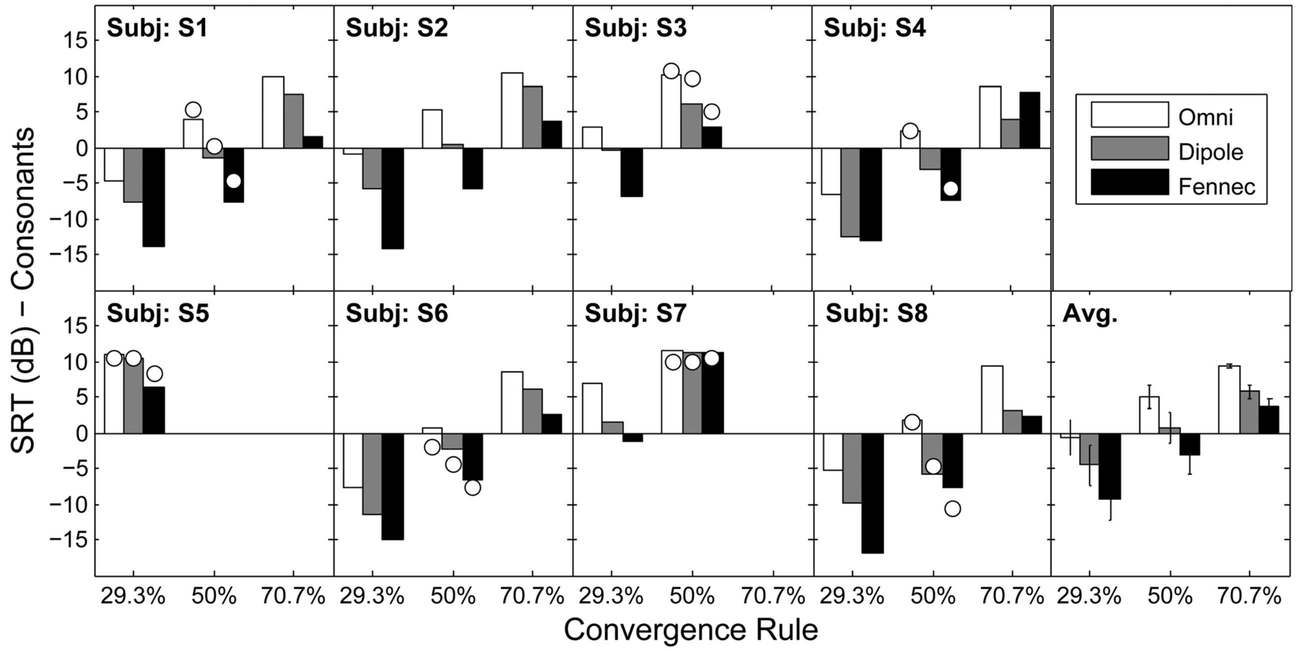

FIG. 2. SRTs measured using consonant materials are plotted for each subject (individual panels) in each condition. Open circles indicate validation SRTs measured after the primary test. The figure panel labeled "Avg." plots the across-subject means with error bars indicating standard error of the means.

and S7 were not able to achieve a 70.7% correct SRT (even at the maximum possible SNR of 12 dB) for the consonant stimuli while Subject S5 was able to achieve neither a 50% nor a 70.7% correct SRT for both the consonant and vowel databases. The results of the second, validity-check SRT measurements are plotted as open circles. Lower SRT values indicate better performance.

In general, the three adaptive convergence criteria yielded percent-correct performance that agreed with the targets of 29.3%, 50%, and 70.7% correct. The percentages of correct response were calculated for each convergence criterion, for each subject, for each processing type, and for each target stimulus set. Averaged across subject and processing type, the measured percent correct scores for the consonant stimulus set were 29.3% [standard deviation (s.d.) of 1.7%], 50.0% (s.d. of 2.1%), and 70.3% (s.d. of 1.9%), for the 29.3%, 50%, and 70.7% convergence criteria, respectively. The measured percent correct scores for the vowel stimulus set were 30.7% (s.d. of 2.7%), 51.9% (s.d. of 2.2%), and 70.7% (s.d. of 3.0%), for the 29.3%, 50%, and 70.7% convergence criteria, respectively.

Averaged across subject and processing type, the SRT required to achieve 29.3%, 50%, and 70.7% correct were −4.8, 0.9, and 6.2 dB, respectively, for the consonant stimuli

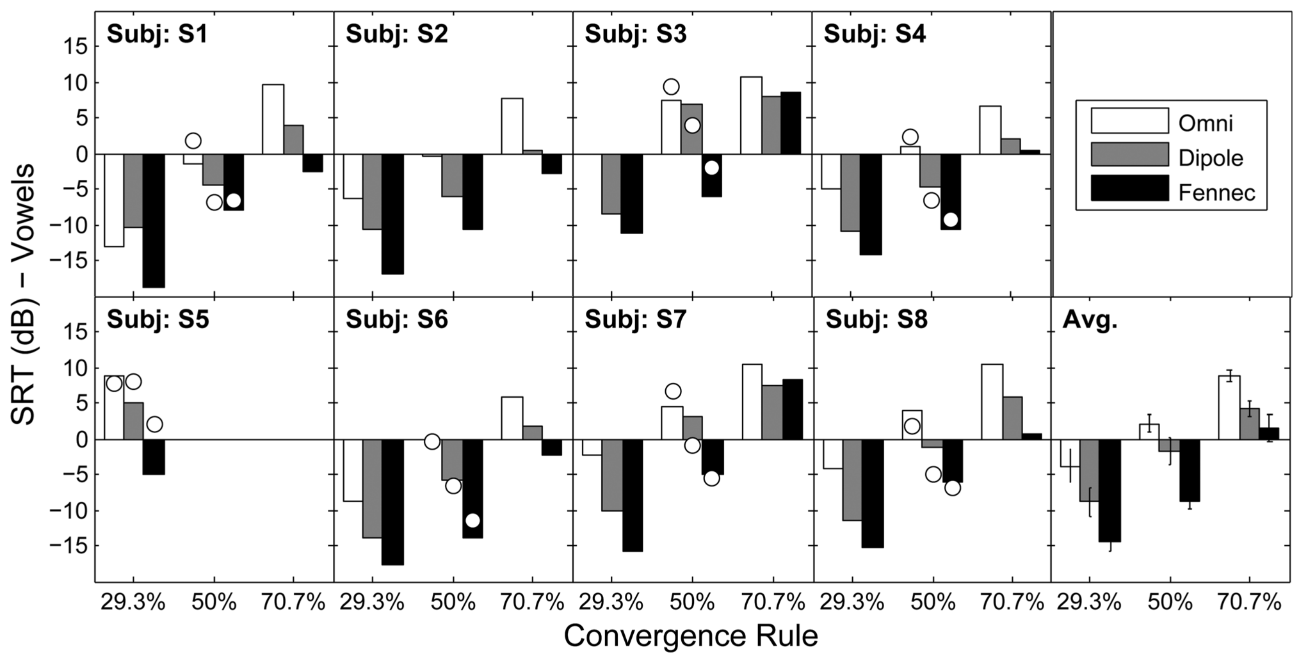

FIG. 3. SRTs measured using vowel materials are plotted for each subject (individual panels) in each condition. Open circles indicate validation SRTs measured after the primary test. The figure panel labeled "Avg." plots the across-subject means with error bars indicating standard error of the means.

and −9.0, −2.8, and 4.8 dB, respectively, for the vowel stimuli. This trend for increasing in SRT with increasing target percent-correct is expected. Averaged across subject and convergence criterion, the SRT attained with Omni, Dipole, and Fennec processing was 4.6, 0.6, and −2.9 dB, respectively, for the consonant stimuli and 2.3, −2.1, and −7.2 dB, respectively, for the vowel stimuli. This indicates that, on average, Fennec outperformed Dipole which outperformed Omni processing.

Figure 4 shows the consonant (upper panel) and vowel (lower panel) SRT benefit for Dipole and Fennec processing over Omni processing averaged over subject for each convergence criterion. These numbers were generated by subtracting the Dipole and Fennec SRTs from the Omni SRTs, and then averaging across all subjects. These SRT benefits ranged from 3.6 to 5.0 dB for Dipole processing and 5.8 to 10.7 dB for Fennec processing. The SRT benefits of Fennec over Dipole processing (calculated in a similar manner and not shown in Fig. 4) are 2.2 to 7.0 dB. Averaged across all three convergence criteria, the SRT benefits of Dipole and Fennec over Omni processing were 4.0 and 7.5 dB, respectively, for the consonant stimuli and 4.4 and 9.5 dB, respectively, for the vowel stimuli. Similarly, the average SRT benefits of Fennec over Dipole processing were 3.6 dB for consonant and 5.1 dB for vowel stimuli.

Six one-way repeated-measures analyses of variance (ANOVAs) were calculated to quantify the effect of processing type (the independent variable) with the subject treated as a dependent variable. Six individual ANOVAs were calculated for each combination of stimulus type (consonant and vowel) and convergence criterion (29.3%, 50%, and 70.7% correct). The ANOVA results are summarized in Table II. Processing type was found to be significant ($p < 0.01$) for all conditions tested. Significant pairwise differences were calculated using *post hoc* Scheffé tests which determined that, for almost all conditions, Fennec outperformed Dipole which outperformed Omni processing. The only exceptions were that the performance differences between Fennec and Dipole processing and between Dipole and Omni processing did not reach significance for the consonant materials tested using the 70.7% convergence rule (please note that only 5 of the 8 subjects were tested on this condition).

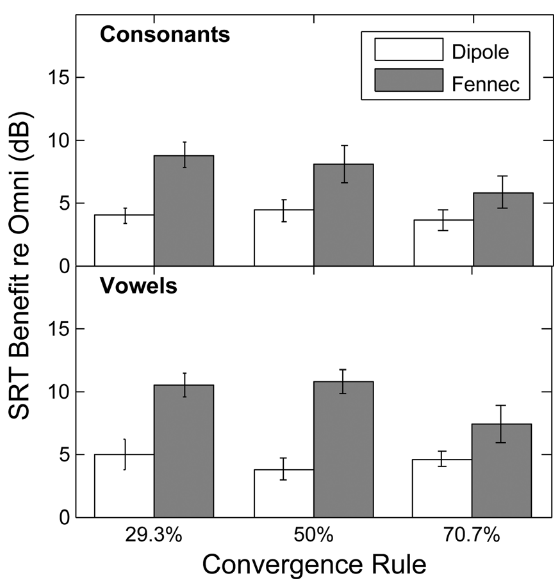

FIG. 4. Average SRT benefit provided by the Dipole and Fennec processing compared to Omni processing for each convergence rule tested. Error bars plot the standard error of the mean across subjects.

A final, one-way repeated-measures ANOVA was performed to compare the validity-check SRTs (shown by open circles in Figs. 2 and 3) with the corresponding experimental SRT scores (shown by the bars in Figs. 2 and 3) with subject, target stimulus set, and processing type considered as dependent variables. No significant difference was found $[F(1, 71) = 1.27,\ p = 0.69]$.

IV. DISCUSSION

The 3.6 to 5.0 dB SRT benefits of Dipole over Omni processing observed in the current study are higher than other directional-microphone SRT benefits previously reported in the literature for CI listeners. For example, Chung *et al*. (2006) reported a 3.5 dB benefit for a hypercardioid directional over an omni-directional microphone. One source of the higher Dipole relative to Omni SRT benefit observed here might be the particular interaction of noise source location with the dipole-microphone directional response. Chung *et al*. (2006) evaluated beamformers using a noise field generated by playing uncorrelated noise from 8 speakers at 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°. In this case, no particular loudspeaker was positioned exactly at either of the theoretical free-field hypercardioid azimuth-plane null locations (120° and 240°). In the current study, on the other hand, two of the three interference sources were positioned at 90° and 270°, which correspond exactly to the theoretical free-field azimuth-plan nulls of a dipole directional microphone. This "favorable" positioning of interference sources may have contributed to the slightly elevated SRT benefits for this particular configuration. It should be noted, however, the combination of mounting the microphones on a head and presenting sources in reverberant conditions minimize the possible benefits that might arise from this interaction.

The 2.2 to 7.0 dB SRT benefit of Fennec over Dipole processing in a moderately-reverberant environment ($T_{60} = 350$ ms) compares favorably to the SRT benefit for the related processing technique of Yousefian and Loizou (2013), which uses estimated input coherence to estimate the input SNR which is then processed through a Wiener filter to yield the particular time/frequency attenuation terms [analogous to those generated by Eq. (3) for Fennec]. Specifically, they reported SRT benefits relative to a directional microphone of 2 to 7 dB in a mildly-reverberant room ($T_{60} = 220$ ms) and of 0 to 2 dB in a more reverberant room ($T_{60} = 465$ ms). The results from the current study, with reverberation conditions intermediate between these two, cover a range almost identical to their low-reverberation results. This suggests that Fennec may be less susceptible to reverberation than their processing. One possible contributing factor to the apparent sensitivity of their processing to reverberation relates to the fact that it is based

TABLE II. Summary of one-way repeated measures ANOVAs with a main factor of processing type conducted for individual subjects for the two stimulus sets and for each convergence rule. Significant pairwise differences (as indicated by *post-hoc* Scheffe tests at the 0.05 level) are also shown.

Stimulus set	Convergence rule	d.f.	F	p	Significant differences (*post-hoc* Scheffé at 0.05)
Consonants	**29.3%**	2, 14	51.77	<0.0001	Omni > Dipole > Fennec
	50%	2, 12	26.89	<0.0001	Omni > Dipole > Fennec
	70.7%	2, 8	9.70	0.007	Omni > Fennec
Vowels	**29.3%**	2, 14	50.13	<0.0001	Omni > Dipole > Fennec
	50%	2, 12	58.61	0.0003	Omni > Dipole > Fennec
	70.7%	2, 12	22.60	<0.0001	Omni > Dipole = Fennec

upon estimated coherence and an underlying assumption that high SNR portions of the input (dominated by the target speech) result in high coherence between microphone signals. Coherence becomes more difficult to estimate accurately as the reverberation level increases—particularly when estimated using finite-length FFTs—as done in their technique. The Fennec strategy, on the other hand, is based solely on inter-microphone phase differences, and preliminary testing by Goldsworthy (2005) and Goldsworthy *et al.* (2009) indicated that such phase differences are less sensitive to reverberation. Another factor that might contribute to the sensitivity of their processing to reverberation is the fact that they employ a Wiener filter to translate estimated SNR into the specific time/frequency attenuation terms applied to the signal. Although such a filter may reduce processing artifacts evident in the output signal, the least mean-square-error criteria used in Wiener filters may adapt too slowly in the presence of reverberation. Goldsworthy (2005) argued that noise reduction for CI users can be implemented more aggressively, since they are relatively insensitive to processing artifacts. With a fast-acting time constant of only 10 ms, Fennec takes such an aggressive approach.

Hersbach *et al.* (2013) evaluated another related approach that uses forward and backward pointing directional microphones to estimate SNR and generate time/frequency attenuation analogous to Eq. (3). They report an average SRT benefit of 4.6 dB relative to the adaptive BEAM™ strategy of Spriet *et al.* (2007). This average SRT benefit falls almost at the mid-point of the range of Fennec SRT benefits over Dipole from the current study. It is difficult to compare these numbers directly, however, due to the difference in reference comparators (adaptive BEAM™ for their study vs non-adaptive Dipole for the current study). Further, their evaluation was conducted in a low-reverberation environment, which was unlike the moderately-reverberant environment from the current study. It should be expected that the performance of their system would deteriorate in the presence of more reverberation.

The most relevant direct comparison of a traditional null-steering approach to the current study is reported in Spriet *et al.* (2007) and assesses the performance of CI listeners using BEAM™ processing in an environment with three multi-talker babble noise sources at 90°, 180°, and 270°. They found an average SRT benefit (relative to an omni-directional response) of 6.5 dB when the overall noise level was set at 55 dB sound pressure level (SPL) and of 11.6 dB when the overall noise level was set to 65 dB SPL. This is similar to the 5.8 to 10.7 dB SRT benefit for Fennec relative to Omni. One interesting similarity between Fennec and BEAM™ is that both techniques appear to yield a greater SRT benefit over Omni when the noise level is louder (observe the trend in Fig. 4 of higher SRT benefits for lower convergence rules; their noise level tends to be louder).

The performance similarity of BEAM™ and Fennec is interesting, especially since these two adaptive processing methods operate with two different processing goals. BEAM™ operates by adaptively steering nulls toward interference sources in the listening environment, while Fennec operates by adaptively identifying and preserving the time/frequency components of the input where the SNR is high.

Given this distinction, it might be possible to combine the Fennec strategy with null-steering beamforming. Hersbach *et al.* (2013) demonstrated an initial step in this direction by using aspects of spatial-filtering to serve as a post-filter for null-steering beamforming. They demonstrated that such a combination yielded synergistic benefits; however, more sophisticated combinations of null-steering beamforming and spatial-filtering could be implemented. Specifically, the Fennec algorithm estimates angle of incidence for incoming sounds and uses this information to attenuate sounds that are not arriving from straight ahead of the listener. This location information could be supplied as an assist to a subsequent null-steering beamformer stage to help reduce, for example, the undesirable null-steering target cancellation that can arise from non-ideal acoustic conditions such as reverberation. Furthermore, the Fennec strategy could be modified so that rather than applying the spatial attenuation to the dipole output, it is applied to the front and back microphone signals independently. This would allow for a first-pass noise reduction system that could then be fed into a back-end null steering routine. There are numerous other combinations of null-steering beamforming and spectral-based spatial-filtering that may yield synergistic benefits. The key processing advantage of the Fennec over null-steering strategies is that it uses a relatively fast spectral analysis allowing the strategy to adapt quickly to changing acoustic environments.

It is possible that the Fennec algorithm would also provide speech-reception benefits for hearing-aid users. However, the implementation in the present study was configured to be relatively aggressive in terms of spatial filtering; consequently, the algorithm introduced processing artifacts that normal-hearing listeners could perceive. Optimizing the algorithm for hearing-aid users might require a less aggressive implementation to reduce processing artifacts, which might ultimately affect the benefit derived from the algorithm. Studies are required to determine such an optimal trade-off between spatial filtering and perceptual quality specifically for hearing-aid users.

V. CONCLUSION

The presented study evaluated a two-microphone spatial-filtering algorithm in a moderately reverberant environment ($T_{60} = 350$ ms) using consonant and vowel-identification measures with three simultaneous noise sources (time-reversed sentences) located at 90°, 180°, and 270°. The spatial-filtering algorithm provided 5.8 to 10.7 dB benefit over an omni-directional response. These results demonstrate that a conceptually-simple spatial-filtering algorithm that operates on two closely-spaced microphones in a BTE configuration can yield substantial speech-reception-in-noise improvements for CI listeners.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. R01 DC 000117, R01 DC DC7152, and 2R44DC010524-02 from the National Institutes of Health, NIDCD.

Burkhard, M. D., and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**(1), 214–222.

Chung, K., and Zeng, F.-G. (**2009**). "Using hearing aid adaptive directional microphones to enhance cochlear implant performance," Hear. Res. **250**, 27–37.

Chung, K., Zeng, F.-G., and Acker, K. N. (**2006**). "Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance," J. Acoust. Soc. Am. **120**(4), 2216–2227.

Chung, K., Zeng, F.-G., and Waltzman, S. (**2004**). "Using hearing aid directional microphones and noise reduction algorithms to enhance cochlear implant performance," ARLO **5**(2), 56–61.

Colburn, H. S. (**1994**). "Computational models of binaural processing," in *Auditory Computation*, edited by H. Hawkins and T. McMullin (Springer-Verlag, New York).

Dawson, P. W., Mauger, S. J., and Hersbach, A. A. (**2011**). "Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus® cochlear implant recipients," Ear Hear. **32**(3), 382–390.

Frost, O. L. (**1972**). "An algorithm for linearly constrained adaptive array processing," Proc. IEEE **60**(8), 926–935.

Fu, Q.-J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," NeuroReport **13**(13), 1635–1639.

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing," J. Acoust. Soc. Am. **104**(6), 3586–3596.

Goldsworthy, R. L. (**2005**). "Noise reduction algorithms and performance metrics for improving speech reception in noise by cochlear implant users," Doctoral thesis, Harvard-MIT Division of Health Sciences and Technology.

Goldsworthy, R. L., Delhorne, L. A., Braida, L. D., and Reed, C. M. (**2013**). "Psychoacoustic and phoneme identification measures in cochlear-implant and normal hearing listeners," Trends Amplification **17**(1), 27–44.

Goldsworthy, R. L., Desloge, J. D., and Zurek, P. M. (**2009**). "Evaluation of a novel noise reduction method," *Conference on Implantable Auditory Prosthesis*, Asilomar, CA.

Greenberg, J. E., and Zurek, P. M. (**1992**). "Evaluation of an adaptive beamforming method for hearing aids," J. Acoust. Soc. Am. **91**(3), 1662–1676.

Griffiths, L. J., and Jim, C. W. (**1982**). "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propag. **30**, 27–34.

Hamacher, V., Doering, W. H., Mauer, G., Fleishmann, H., and Hennecke, J. (**1997**). "Evaluation of noise reduction systems for cochlear implant users in different acoustic environments," Am. J. Otol. **18**(6), S46–S49.

Hanekom, J. J., and Shannon, R. V. (**1998**). "Gap detection as a measure of electrode interaction in cochlear implants," J. Acoust. Soc. Am. **104**(4), 2372–2384.

Henry, B. A., and Turner, C. W. (**2003**). "The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners," J. Acoust. Soc. Am. **113**(5), 2861–2873.

Hersbach, A. A., Arora, K., Mauger, S. J., and Dawson, P. W. (**2012**). "Combining directional microphone and single-channel noise reduction algorithms: A clinical evaluation in difficult listening conditions with cochlear implant users," Ear Hear. **33**(4), e13–e23.

Hersbach, A. A., Grayden, D. B., Fallon, J. B., and McDermott, H. J. (**2013**). "A beamformer post-filter for cochlear implant noise reduction," J. Acoust. Soc. Am. **133**(4), 2412–2420.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5 Pt. 1), 3099–3111.

Hochberg, I., Boothroyd, A., Weiss, M., and Hellman, S. (**1992**). "Effects of noise and noise suppression on speech perception by cochlear implant users," Ear Hear. **13**(4), 263–271.

Hu, Y., and Loizou, P. C. (**2008**). "A new sound coding strategy for suppressing noise in cochlear implants," J. Acoust. Soc. Am. **124**(1), 498–509.

Hu, Y., and Loizou, P. C. (**2010**). "Environment-specific noise suppression for improved speech intelligibility by cochlear implant users," J. Acoust. Soc. Am. **127**(6), 3689–3695.

Hughes, M. L., and Goulson, A. M. (**2011**). "Electrically evoked compound action potential measures for virtual channels versus physical electrodes," Ear Hear. **32**(3), 323–330.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **AU-17**, 225–246.

Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**(1), 35–39.

Kokkinakis, K., and Loizou, P. (**2008**). "Using blind source separation techniques to improve speech recognition in bilateral cochlear implant patients," J. Acoust. Soc. Am. **123**, 2379–2390.

Kollmeier, B., and Koch, R. (**1994**). "Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction," J. Acoust. Soc. Am. **95**(3), 1593–1602.

Kollmeier, B., Peissig, J., and Hohmann, V. (**1993**). "Real-time multiband dynamic compression and noise reduction for binaural hearing aids," J. Rehabil. Res. Dev. **10**(1), 82–94.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2), Supplement 2:467+.

Lim, J. S., and Oppenheim, A. V. (**1979**). "Enhancement and bandwidth compression of noisy speech," Proc. IEEE **67**(12), 1586–1604.

Liu, C., and Eddins, D. A. (**2012**). "Measurement of stop consonant identification using adaptive tracking procedures," J. Acoust. Soc. Am. **132**(3), EL250–EL256.

Mackie, K., and Dermody, P. (**1986**). "Use of monosyllabic adaptive speech with young children," J. Speech Hear. Res. **29**, 275–281.

Margo, V., Schweitzer, C., and Feinman, G. (**1997**). "Comparisons of Spectra 22 performance in noise with and without an additional noise reduction preprocessor," Semin Hear **18**(4), 405–415.

Morita, T., Naito, Y., Tsuji, J., Nakamura, T., Yamaguchi, S., and Ito, J. (**2004**). "Relationship between cochlear implant outcome and the diameter of the cochlear nerve depicted on MRI," Acta Otolaryngol. **124**(s551), 56–59.

Muchnik, C. (**1994**). "Auditory temporal resolution and open speech recognition in cochlear implant recipients," Scand. Aud. **23**, 105–109.

Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**(2), 961–968.

Rife, D. D., and Vanderkooy, J. (**1989**). "Transfer-function measurement with maximum-length sequences," J. Aud. Eng. Soc. **37**(6), 419–444.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**(6), L71–L74.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature **416**(6876), 87–90.

Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., van Wieringen, A., and Wouters, J. (**2007**). "Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the Nucleus Freedom Cochlear Implant System," Ear Hear. **28**(1), 62–72.

Stadler, R. W., and Rabinowitz, W. M. (**1993**). "On the potential of fixed arrays for hearing aids," J. Acoust. Soc. Am. **94**(3), 1332–1342.

Tecca, J. E., and Binnie, C. A. (**1982**). "The application of an adaptive procedure to the California consonant test for hearing aid evaluation," Ear Hear. **3**(2), 72–76.

Vanderkooy, J. (**1994**). "Aspects of MLS measuring systems," J. Aud. Soc. Am. **42**(4), 219–231.

Van Hoesel, R. J., and Clark, G. M. (**1995**). "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," J. Acoust. Soc. Am. **97**(4), 2498–2503.

Wall, L. G., Fox, A., and Davidson, A. (**1989**). "Presentation level determination for CV tokens using a computer-controlled adaptive procedure," J. Commun. Disord. **22**, 193–207.

Weiss, M. R. (**1993**). "Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor," J. Rehab. Res. Dev. **30**(1), 117–128.

Widrow, B., Glover, J. R., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, E., and Goodlin, R. C. (**1975**). "Adaptive noise cancelling: Principles and applications," Proc. IEEE **63**, 1692–1716.

Wouters, J., and Vanden Berghe, J. (**2001**). "Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system," Ear Hear. **22**, 420–430.

Yang, L.-P., and Fu, Q.-J. (**2005**). "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," J. Acoust. Soc. Am. **117**(3), 1001–1004.

Yousefian, N., Kokkinakis, K., and Loizou, P. C. (**2010**). "A coherence-based algorithm for noise reduction in dual-microphone applications," in *Proceedings of the European Signal Processing Conference (EUSIPCO'10)*, Alborg, Denmark, pp. 1904–1908.

Yousefian, N., and Loizou, P. C. (**2012**). "A dual-microphone speech enhancement algorithm based on the coherence function," IEEE Trans. Audio, Speech, Language Process. **20**(2), 599–609.

Yousefian, N., and Loizou, P. C. (**2013**). "A dual-microphone algorithm that can cope with competing-talker scenarios," IEEE Trans. Audio, Speech, Language Process. **21**(1), 145–155.

Zeng, F.-G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J., and Segel, P. (**2002**). "Speech dynamic range and its effect on cochlear implant performance," J. Acoust. Soc. Am. **111**(1), 377–386.